Review of *Causation in Science*, by Yemima Ben-Menahem

Princeton: Princeton University Press, 2018. pp. xiii + 201

Matt Farr

University of Cambridge

hello@mattfarr.co.uk • mattfarr.co.uk

Forthcoming in *Mind*

September 23, 2021

1 Introduction

Whatever you think of Bertrand Russell's famous claim that the 'law of causality' is (at least as of 1912) redundant in the 'advanced sciences', it is nonetheless the case that a variety of concepts tangled up with the idea of causation — such as determinism and locality — remain ubiquitous within physics and elsewhere across the sciences. Yemima Ben-Menahem's excellent book, *Causation in Science*, focuses not on reductive metaphysical accounts of these notions but instead the roles they play within physics, and the relationships that hold between them, through a series of detailed historical case studies. The book is a highly-informed, illuminating and detailed analysis of these causal notions central to scientific practice, with many highly novel discussions about their nuanced relationships, and as such it adds significantly towards the philosophy of causation and philosophy of science more generally, and also sets out a number of new approaches for philosophers of science to explore regarding the nature of causality and causal reasoning. In this review, I set out the main aims and themes of the book, assess whether the types of physical constraints explored by the book deserve the name 'causal', and discuss what really hangs on the distinction between causal and non-causal constraints in physics.

2 Aims of the book

So what *is* causation? Clearly it's a much-used term within science, but there are many different concepts that have come to be called 'causal' across the sciences for varying reasons rarely set out or justified by the authors of published scientific research. Ben-Menahem takes on an important and difficult task: to disentangle various instances of a particular class of causal notions in fundamental science — *constraints* — and to trace the relationships between them. Among these, the book provides analyses of the following: determinism; stability; locality; symmetry; conservation laws; and least action principles. Finally the topic of causal reductionism is itself tackled in the concluding chapter. As Ben-Menahem notes, each of these chapters can be taken in isolation from the others, but among the complex issues covered by the individual chapters there are several background themes tying the book together: (1) in order to understand the role of causality in science, philosophers of science should focus more on the concept of causal constraints and less on causal relations between events; (2) causality is a 'cluster concept' that ranges over a number of interrelated concepts within science; (3) reductionist accounts of causation in science are largely naïve and insufficiently motivated by scientific practice.

2.1 Causal Constraints

Causality can lay claim to being the most ambiguous term in science, referring to a wide range of tenuously related notions, and not clearly to a single concept. It can be used to refer to global features of the structure of spacetime or to the local nature of interactions and mechanisms, to a fundamental feature of subatomic interactions or to an emergent statistical feature of macroscopic systems. As a result, many have suggested that we should simply do away with the idea of causality in physics altogether. Causal notions have been criticised as too vague and imprecise for science, and as such something to be eradicated. For instance, Gustav Kirchhoff (1876, p. v) deemed the search for causes in physics as undesirable on the grounds of causal notions being 'infected by the vagueness from which the notions of cause of striving cannot be freed',[1] and Ernst Mach (1906, p. 278) noted that causal notions 'lack precision'[2] and for this reason causality plays an increasingly minimal role in physics. Others such as Russell and John Norton (2007) have deemed causality an outdated notion no longer in keeping with the notions central to fundamental physics.

Various attempts have been made to rehabilitate causality through providing rigorous analyses of causal relations and causal structure, such as with the axiomatic accounts of causal modelling of Pearl (2000) and Spirtes et al. (2001). But Ben-Menahem takes a key aspect of causality in science to be overlooked by philosophers, namely the role of causal constraints in science. The book, she notes, 'seeks to shift our attention from causal relations between individual events [...] to the more general causal constraints found in science, and the relations between them,' (p. 1) noting that focus on scientific practice 'led me to the conclusion that the notion of causal constraint is far more germane than that of causal relations between individual events' (p. ix). The aim is not to simply shift the focus of philosophers of causation from causal relations to causal constraints, but rather to 'suggest a broader perspective on causation and a new research program for the philosophy of causation' (p. 1).

2.2 Causal Pluralism

In response to causal sceptics and causal eliminativists like Mach, Russell and Norton, Ben-Menahem moves away from a singular notion of causation and a singular account of what causal things have in common: 'I have relinquished

[1]Translation by Frisch (2014, p. 2).
[2]Translation by Frisch (2014, p. 4).

the search for *the* definition of causation, instead taking causation to be a cluster concept comprising a broad range of causal notions' (p. 9). Instead, she says, the argument of the book is that 'as soon as we shift our attention from the familiar paradigms of breaking a glass or tickling a baby to determinism, locality, stability and conservation laws, it becomes evident that causal notions permeate fundamental science' (p. 9). There is nothing, she suggests, that 'is both necessary and sufficient' that could constitute a 'definition of the notion of cause' that accounts for the variety of causal notions used in fundamental science, nor some underlying causal principle used within science. Instead, 'I have therefore relinquished the quest for such a definition and adopted a pluralistic approach, taking the notion of cause to be an irreducible cluster concept covering various constraints imposed but the theories we employ' (p. 12).

2.3 Causal Antireductionism

Through its focus on scientific practice, its pluralistic analysis of causal constraints, and its resulting move away from focusing on causal relations between events, *Causation in Science* fits with a trend of philosophical analyses of causation that avoid overly metaphysical analysis. Ben-Menahem stresses she does not 'aspire to completeness' in her account of causation, nor does she take the kinds of causal notions discussed to 'have an a priori basis' (p. x), noting that even when such notions stem from 'deeply rooted intuitions, they are part of science...[and are] always subject to reevaluation' (ibid). As such, her analysis does not amount to a polemical defence of some particular philosophical theory of causation. But this is not to say that there are no underlying targets of her approach. The preface notes that '[i]n drawing attention to the spectrum of causal constraints that guide fundamental science, the argument set forth in this book takes issue with causal eliminativism [...and in] elucidating the structure of inter theoretic relations, it challenges causal reductionism' (p. xi).

So what we have here is a fascinating book that provides a detailed and thoroughgoing analysis of the complex relationships between causal constraints within science and paints a far more nuanced account of causation than that assumed by causal eliminativists and causal reductionists. Insofar as there is a central conclusion of this book it is that philosophers have a tendency to try to fit the world into categories into which it simply does not fit, and that by focusing on the usage of putative causal notions in scientific practice we can gain nontrivial knowledge about the interplay of notions like determinism, stability, locality and symmetry, and so better understand the role of causality in science. Pulling these themes together, the final chapter of the book is ultimately the closest thing to a polemical take on the philosophy of causation, showing the various pitfalls that arise from trying to view causation in science through a reductionist lens.

3 Summary of chapters

After setting out the aims and motivations of the book in chapter 1, chapters 2 and 3 discuss the relationship between the notions of determinism and stability. Alongside a detailed and informative historical overview of different notions of determinism, Ben-Menahem uses these chapters to explicate what she takes to be a 'neglected causal notion', *stability*, which she describes as 'a distinct causal category, independent of the category of determinism with which it is often conflated' (p. 60). Stability is introduced as a property of states to which systems return after 'having been subject to small changes', and is a widespread notion within physics, particularly in statistical mechanics (the main topic of chapter 3) when dealing with things like equilibrium states. Stability is causal in Ben-Menahem's terms precisely because it concerns the notion of change over time — the property of stability acts as a constraint on how much a system changes in certain physical interactions.

As Ben-Menahem stresses, determinism is a notion that has changed from being generally understood in causal terms, such as that the same causes lead to same effects, to noncausal terms, such as that for any particular state of the world, there is only one possible history compatible with the laws (something that led Russell to the conclusion that the determinism of laws of physics is incompatible with a causal interpretation). Stability, Ben-Menahem argues, is a more clearly causal notion, since it concerns how perturbations affect systems, and system's tendency to return or fail to return to some particular state given a certain kind of influence. An ongoing theme through chapters 2 and 3 is that the distinction between stable and unstable is more useful within science than the distinction between necessity and contingency. Thinking in terms of whether some state arises out of necessity or contingency is inextricably tangled up with philosophical worries about fatalism, teleology and directionality of processes. Addressing these related issues, Ben-Menahem provides engaging discussions of how the concept of stability plays a role in clearing up such worries, such as in apparently goal-oriented or teleological explanations in evolutionary biology (ch 2), where we see how even contingent processes lead predictably towards some stable end (such as the evolution of eyes or wings), and the appearance of directionality in statistical mechanics (ch 3) in the evolution of systems from unstable low-entropy states to stable high-entropy states.

Chapter 4 takes on the relationship between determinism and locality, something 'all but ignored' within the philosophical literature (p. 83). The chapter largely focuses on the non-local nature of quantum mechanics and its relationship to relativistic causality (i.e. the kind of locality imposed by special relativity theory). In analysing the various ways in which determinism and locality impinge upon each other, this chapter produces many important insights into the causal nature of quantum mechanics. For instance, Ben-Menahem notes that the regularity of the non-local correlations described by quantum mechanics emphasises the deterministic nature of quantum mechanics: 'surprisingly then a grain of determinism turns out to be necessary, de facto, for nonlocality; it is necessary for the formulation of a nonlocal theory' (p. 85). As such, (non)locality and (in)determinism are not fully independent concepts. Indeed, as is well known, the fact that quantum mechanics is both non-local and yet abides by no-signaling principles, means that the theory has an intriguingly complex causal structure.

We can better understand this, Ben-Menahem suggests, by focusing on the 'intricate dance between indeterminism and locality' (p. 89) performed by the theory: 'the satisfaction of the no-signaling constraint is made possible by the fact that QM also accommodates (at least a measure of) indeterminism', with the combination of entanglement and no-signaling 'strik[ing] a delicate balance between determinism and indeterminism' (p. 90). The chapter goes on in detail considering the balance of determinism and indeterminism in different approaches and interpretations of quantum mechanics.

In Chapter 5, the focus is shifted to the relationship between symmetries and conservation laws. Symmetries, Ben-Menahem contends, are causal insofar as they 'constrain physical change and physical possibility' (p. 110). This marks an interesting take on the use of the term 'causal', since symmetry principles are often appealed to as providing *non*-causal explanations of processes, an issue I'll come back to in more detail below. But the idea here is that since symmetry principles are routinely appealed to in explanations of why certain systems evolved some way and not another, and why certain systems *could* evolve one way but not another, symmetries are sufficiently connected to the idea of causal constraints on the evolution of systems as to be causal. Ben-Menahem demonstrates this in the case of the Pauli Exclusion Principle (PEP) in an engaging discussion (pp. 111–115) arguing for the causal role of PEP in the collapse of white dwarfs in the work of Chandrasekhar, and additionally making the case that PEP is 'directly tied to causal considerations' since it assumes relativistic locality, ultimately arguing that PEP is 'an excellent illustration of the causal function of symmetries in physics' (p. 115).

The chapter continues with further reflection on the very meaning of causality through its discussion of the causal nature of conservation principles. For instance, Ben-Menahem notes how Niels Bohr and Émile Meyerson understood causality in terms of conservation, with Bohr using the term 'causality' to refer to the conservation of momentum and energy, and Meyerson poetically noting that 'the external world [...] appears to us as infinitely changing [...] yet the principle of causality postulates the contrary[, ...] change is only apparent; it must necessarily disclose an identity which alone is real' (Meyerson, 1930, p. 92; quoted by Ben-Menahem on p. 122). Continuing the consideration of symmetries, the chapter ends with a discussion of Curie's principle, which Ben-Menahem acknowledges has been criticised on grounds of being tautologous and so not obviously causal, but responds that understood as a general constraint on change, Curie's principle 'continues to inform the search for a unified picture of the world' (p. 133).

Chapter 6 presents the final case study of causal constraints, focusing on the principle of least action and its relationship to teleology, a theme previously touched upon in chapter 2. For instance, Ben-Menahem notes that Fermat's principle, a precursor of the principle of least action, 'appears tantalisingly teleological' (p. 142), asking: 'how can light pick out the right path other than by some sort of calculation that takes the entire path into account? The very notion of a *path*, involving special end points, seems to presuppose goal directness. Indeed, the principle was criticised straightaway on account of its ostensible ascription of foreknowledge to nature, in violation of scientific standards' (Ibid). The discussion of this chapter is helpfully placed in the context of discussions of teleology in the work of Spinoza, Descartes, Leibniz and Newton, noting the interplay of causal and teleological modes of explanation with theological views. For instance, Ben-Menahem notes how Newton's own theology supported the idea that God chose a causal over teleological world, as one depending upon initial rather than final conditions (pp. 139-140). Through the historical and practice-led approach, this chapter offers an extremely useful presentation of the least action principle and its role in scientific explanations, and it might have been interesting to see additional engagement with some of the more recent philosophical work on the principle of least action, and relationships with retrocausal and time-symmetric formulations of quantum mechanics.

As noted, the final chapter brings the discussion back to traditional philosophical conceptions of causation with a critique of 'higher-level eliminativism' — the idea that causality is a fundamental or underlying feature of the world and that all causal facts are to be found at the fundamental level, with apparent higher-level causal facts are either reducible to facts about the fundamental level or else false. Ben-Menahem notes this to be the converse position to 'causal republicanism', the view defended by Price and Corry (2007) and others, which holds that while causality is to be eliminated from fundamental physics, causal notions are nonetheless useful in higher-level and everyday contexts. As Ben-Menahem presents republicanism, the idea is that there are higher-level but not fundamental causal facts. While much of the earlier parts of the book can be considered a response to republicans (and also the other kind of fundamental causal eliminativists described by Price & Corry, the causal anarchists), this final chapter concerns itself solely with higher-level eliminativism, focusing primarily with problems concerning causal reduction and emergence. Ben-Menahem rounds the book off with an intriguing discussion of the idea of lawlessness in physics, and how the notion of lawlessness relates to our sense of human freedom.

4 Discussion: 'Causal' vs 'noncausal' constraints; What's in a name?

There has been much work in the past decade in philosophy of science on the apparent role of noncausal explanations in science (see for instance Lange (2016), Reutlinger (2017) and Reutlinger and Saatsi (2018)). This is a literature that receives little discussion in the book despite the overlapping themes. In particular, Lange (2011, 2013, 2016, 2018) refers to 'explanations by constraint' as *non*causal explanations; for Lange, explanations by constraint exhibit explanations of facts that are modally stronger than causal explanations (in Lange's words, they provide a 'a stronger variety of necessity' (2018, p. 27)). In other words, the relevant explanandum is constrained to be some particular way by a non-causal fact, such as the holding of a particular symmetry principle:

> [S]patiotemporal symmetries and the principle of relativity do not describe causal relations [..., r]ather, they impose constraints on the laws of nature; they require that the laws of nature take a certain form. [...] Similarly, the fact that the spacetime interval is invariant is not a matter of the world's network of causal relations. (Lange, 2013, p. 494)

But how does Ben-Menahem's account of constraints as causal relate to Lange's? The parallels with Lange's account of explanations by constraint is acknowledged on pp. 15-16 of the book, though not at great length. As Ben-Menahem notes, one key difference here is terminological: 'whereas the constraints I speak of are *causal*, [Lange] deems explanation by constraint noncausal, implying that the notions of cause and constraint exclude each other' (p. 15). Ben-Menahem relates this to her wider pluralism about causality, noting that '[t]he inclusive concept of causation as comprising *any* constraint on change, regardless of its place in the hierarchy of laws, affords a better understanding of the function of causal notions in science' (p. 16). This picks on the key rule of thumb in Ben-Menahem's notion of causation — for something to be regarded as causal, it must be associated in the right way with time and change.

> I understand causal constraints as general constraints on change. As such they constitute the conceptual scaffolding of the natural sciences, and differ from purely mathematical constraints, which are indifferent to temporal change and evolution. (p. x)

Such a condition is hard to make strict since many of the key cases in the debate about non-causal explanations in science involve things that are atypical with respect to time and change. For example, consistency constraints in universes containing closed timelike curves are often appealed to as non-causal explanations for why someone in such a world is unable to kill their younger self, despite clearly associated with time and change. Given the complexity of the relationship between global consistency constraints such as this and temporal change and evolution, Ben-Menahem's rule of thumb is ultimately hard to apply one way or the other.

Nonetheless, the debate about the role of putatively noncausal explanations in science has been active for several years and the token examples used of such explanations overlap with the kinds of constraints that Ben-Menahem deems 'causal', so further discussion here is merited. One might perhaps think that the kinds of constraints appealed to are causal in nature in that they are ultimately describe the large-scale features of causal processes — after all, symmetry principles hold only insofar as they accurately describe the behaviour of dynamical systems. Within the literature on non-causal explanations, there is a variety of approaches to how causal and non-causal explanations are related. As Reutlinger (2017) notes: *reductivists* take non-causal explanations to be causal explanations in disguise, as just described in the case of symmetry principles; *pluralists* take both causal and non-causal explanations to be autonomous kinds of legitimate scientific explanation; and *monists* take there to be an underlying structure of explanation that both causal and non-causal explanations exhibit. Ben-Menahem's account doesn't neatly fit this set of options, primarily because what is deemed 'causal' within the book appears to be incorporate many kinds of constraint often deemed non-causal.

Returning to the topic of the Pauli exclusion principle, Ben-Menahem notes it as being used as a token 'noncausal' explanation by Railton (1978), Salmon (1989), and Lange (2016). In response, Ben-Menahem notes that the exclusion principle is not properly described as noncausal since it 'does not spring from purely mathematical considerations' (p. 114). She adds '[m]athematics cannot, by itself, account for the existence of particles that are indistinguishable from one another, act in conformity with Fermi-Dirac statistics, and are excluded from occupying the same state' (ibid), instead contending that the principle stems from manifestly causal considerations, such as locality, and moreover has been 'very successfully used as a causal constraint' (p. 115). Here there is the worry that this issue may be terminological in nature — after all, Lange distinguishes non-causal explanations by constraint as distinct from 'purely mathematical' explanations. But what appears key here is the actual use of causal considerations and causal reasoning in the development and establishment of the principle. This is reflective of the general practice-first approach to science endorsed within the book, fitting with the approach of Frisch (2014) by focusing on the causal aspects involved in scientific practice and scientific reasoning when discovering or confirming more abstract principles like the exclusion principle.

Perhaps there's a certain arbitrariness to this — after all, what's in a name? If one takes 'causality' to necessarily refer to the satisfying of certain axioms like locality, screening-off, and so on, then things like global consistency constraints will not be deemed 'causal'. On the other hand, if we take a looser account of what counts as 'causal', and we wish to distinguish constraints that are apparently mathematical from those that are dynamical — such as constraining change over time of objects —, then it is useful to call certain global constraints as 'causal'. Is this then just a terminological matter? And does the apparent arbitrariness or vagueness as to the label 'causal' simply echo the problems raised by the likes of Kirchhoff and Mach and other causal eliminativists who ultimately saw causal terms as responsible for letting too much obscurity into scientific discourse? Ultimately, a key lesson to extract from the book is not to hang too much metaphysical baggage on the notion of causation; to instead treat 'causal' as something of a family resemblance concept, one that is pluralistic and does not obviously admit of some set of necessary and sufficient conditions. It is certain that there is *something* about the kinds of constraints considered in this book, their interrelations, and how they have been used in the development of the physical theories, that plays a special role in how we reason about the world and about how physics implies the world is structured. And through understanding causation in science in this way, the book offers an important new resource.

References

Frisch, M. (2014). *Causal Reasoning in Physics*. Cambridge: Cambridge University Press.

Kirchhoff, G. (1876). *Vorlesungen ueber mathematische Physik: Mechanik*. Teubner.

Lange, M. (2011). Conservation laws in scientific explanations: Constraints or coincidences? *Philosophy of Science 78*(3), 333–352.

Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science 64*(3), 485–511.

Lange, M. (2016). *Because Without Cause: Non-Casual Explanations In Science and Mathematics*. Oxford University Press.

Lange, M. (2018). Reply to my critics: On explanations by constraint. *Metascience 27*(1), 27–36.

Mach, E. (1906). *Erkenntnis und Irrtum: Skizzen zur Psychologie der Forschung*. Barth.

Meyerson, E. (1930). *Identity & Reality*. London: Routledge.

Norton, J. D. (2007). Causation as folk science. In H. Price and R. Corry (Eds.), *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, Chapter 2, pp. 11–44. Oxford: Oxford University Press.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Price, H. and R. Corry (2007). A case for causal republicanism? In R. C. Huw Price (Ed.), *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, pp. 1–10. Oxford University Press.

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of science 45*(2), 206–226.

Reutlinger, A. (2017). Explanation beyond causation? new directions in the philosophy of scientific explanation. *Philosophy Compass 12*(2), e12395.

Reutlinger, A. and J. Saatsi (2018). *Explanation beyond causation: philosophical perspectives on non-causal explanations.* Oxford University Press.

Russell, B. (1912). On the notion of cause. *Proceedings of the Aristotelian Society 13*, pp. 1–26.

Salmon, W. C. (1989). 4 decades of scientific explanation. *Minnesota Studies in the Philosophy of Science 13.*

Spirtes, P., C. N. Glymour, and R. Scheines (2001). *Causation, prediction, and search.* The MIT Press.